

基于购物倾向的商品推荐方案研究

王树西 李安渝

(对外经济贸易大学 北京 100029)

摘要 为了提高商品推荐系统的性能,从理解 B2C 电子商务平台客户的购物倾向角度出发,进行实时的商品推荐。本文总结归纳了 B2C 电子商务平台的主要商品推荐位,以及如何基于商品信息建立商品标签,如何判别商品的相似性、相同性。在上述工作的基础上,重点论述了基于客户购物倾向的实时的商品推荐方案。本文对已有工作进行了深入对比和分析,对提出的方案进行了必要的理论分析和性能评估。并从 XBRL 的技术角度,对商品推荐方案进行了改进。

关键词 购物倾向;商品推荐方案;理论分析;性能评估;XBRL

A Commodity Recommend Scheme Based on Customers' Purchase Intentions

WANG Shu-xi LEE An-yu

(University of International Business and Economics, Beijing 100029, China)

Abstract To improve the performance of the commodity recommendation system, this paper proposes two methods: (1) recommend commodity in real-time; (2) understand customers' purchase intentions. In order to achieve the above goals, we summarize commodity recommendation location, and the label of commodity. Based on the above, we propose a scheme to improve the performance of the commodity recommendation system. We conduct an in-depth analysis on existing work, and give necessary theoretical analysis and performance evaluation for the proposed commodity recommend scheme. From the technical point of view, this method improves the above commodity recommend scheme.

Keywords purchase intention; commodity recommend scheme; theoretical analysis; performance evaluation; XBRL

1 引言

商品推荐系统是高级商务智能平台,也可以看做是一种智能的商务代理,为电子商务平台的顾客提供个性化的决策支持、信息服务、商品选择和推荐。系统基于顾客的购买行为、浏览行为、对商品的评价等信息,学习顾客的兴趣,进行商品匹配,向顾客推荐相似的商品。商品推荐系统根据商品内容信息作出推荐,从关于内容的特征描述的事例中得到顾客的兴趣资料。推荐基于:网站最热卖商品、顾客所处城市、顾客过去的购买行为和购买记录,推测顾客将来可能的购买行为等。在 B2C 电子商务中,商品推荐系统

有着广泛的用途,随着 B2C 电子商务的迅速发展,商品推荐系统越来越受到重视,作用越来越大。几乎所有的大型 B2C 电子商务平台,都不同程度的使用了商品推荐系统。建设一个优秀的商品推荐系统需要很多因素,本文试图从如下两个角度出发,探索提高商品推荐系统性能的途径,试图为商品推荐系统提供一种全新的研发思路:(1)基于客户购物倾向进行商品推荐;(2)向客户实时推荐商品。

本文结构如下:(1)深入分析商品推荐系统的工作流程与系统架构,并提出一系列创新点,特别是总结归纳了 B2C 电子商务平台的主要商品推荐位;(2)基于商品信息建立商品标签;(3)判别商品的相似性、相同性;(4)基于客户购物倾向的商品推荐方案。

2 主流技术与分析

商品推荐系统所采用的主流技术,可以简单归结为:“物以类聚人以群分”,所有相关算法都是围绕着上述策略研发。

所谓的“物以类聚”,是指把顾客感兴趣的同类商品向顾客推荐。例如,顾客对商品 p (如 iPad) 感兴趣(购买或者多次浏览),那么把商品 p 的同类或者相似商品 q (如 iPhone) 推荐给顾客。这种推荐策略,也称为“基于内容的推荐(Content-Based)”,其核心技术在于计算商品之间的相似度,当商品 p 与商品 q 之间的相似度超过一个阈值,就认为商品 p 与商品 q 是同类商品。基于内容的商品推荐,优点是简单高效;缺点是过于粗放,因为推荐的同类商品不一定是顾客需要的。例如,顾客买了一台电视机,此时向顾客推荐电视机已经没有意义,因为这属于大宗商品,大多数顾客往往只需要购买一件。而且,在商品种类、型号很多,而且新商品不断涌现的情况下,计算商品之间的相似度,工作量非常大。

所谓的“人以群分”,是分析电子商务平台已有顾客,把顾客分为不同的类别。如果顾客 a 购买了商品 p ,那么把商品 p 推荐给与顾客 a 同属一个类别的其他顾客。这种推荐策略,也称为“协同过滤(Collaborative Based)”,其核心技术在于对现有顾客进行分类。基于协同过滤的商品推荐,优点是推荐范围广,实际上,目前大多数电子商务平台的推荐系统,都采用系统过滤的方法,但这种方法也存在一些缺点。例如,“相似购买兴趣顾客”构成一个矩阵,对于一个新注册的顾客,因为没有他的购买记录,所以无法准确分析其购买兴趣,也无法将其准确的划入到某一个“相似购买兴趣顾客类”中,这就造成冷启动问题;一般来说,因为顾客较多,造成这个矩阵非常巨大、稀疏,这就造成稀疏矩阵问题;一个顾客属于多个不同的集合,当这些集合中的人购买商品的信息很多的时候,这些商品信息都会推荐给这个顾客,就会刮起一场信息风暴,造成信息噪音^[1]。

2.1 基于内容的推荐技术

简单来说,就是“物以类聚”:喜欢一件商品,类似的商品估计你也喜欢。形式化为:

商品集合 $I = \{a_1, a_2, \dots, a_n\}$, I 的子集构成集合 $S = \{I_1, I_2, \dots, I_m\}$, 并且满足:

$$(1) I_i \neq \Phi \quad (i=1, 2, \dots, k)$$

$$(2) I_i \cap I_j = \Phi \quad (i, j=1, 2, \dots, k \text{ 且 } i \neq j)$$

$$(3) I_1 \cup I_2 \cup \dots \cup I_k = I$$

称集合 $S = \{I_1, I_2, \dots, I_m\}$ 为商品集合 I 的一个划分(Partition)。 $I_i (i=1, 2, \dots, k)$ 是商品集合 I 的一个划分块。

例如,通过商品相似度计算,发现商品 A 和 B 相似度较高。商品推荐系统发现顾客 U 喜欢商品 A ,于是得出结论:顾客 U 很可能对商品 B 感兴趣。于是把商品 B 推荐给顾客 U 。

基于内容的推荐系统,其优点是简单、有效,其缺点是特征提取的能力有限,过分细化。这种方法通常被限制在容易分析内容的商品的推荐,而对于一些较难提取出内容的商品,如音乐 CD、电影等,就不能产生满意的推荐效果。很多推荐系统中,都设置了顾客反馈的环节。顾客反馈的方式,大致分为两种:一种反馈方式是直接对商品进行评价,另外一种反馈方式是向其他顾客推荐这种商品。某些推荐系统还添加了互动的环节。例如,如果有顾客推荐了某种商品,那么顾客之间可以互相交流对这种商品的看法。

基于内容的商品推荐方法,可以形式化为: $u(c, s) = \text{score}(\text{ContentBasedProfile}(c), \text{Content}(s))$ 。其中, c 表示顾客, s 表示商品^[2]。

也可以通过经典的向量余弦夹角方法,计算顾客 c 与商品 s 之间的距离:

$$u(c, s) = \cos(\bar{w}_c, \bar{w}_s) = \sum_{i=1}^K w_{i,c} w_{i,s} / (\sqrt{\sum_{i=1}^K w_{i,c}^2} \sqrt{\sum_{i=1}^K w_{i,s}^2})$$

对于一个顾客,首先找到他常用的标签,然后对于这些常用标签,找到具有这些标签的最热门的商品,推荐给这个顾客。顾客 u 对商品 s 的兴趣可以用如下的公式度量^[3]:

$$p(u, s) = \sum_b n_{u,b} n_{b,s}$$

其中, $B(u)$ 是顾客 u 打过的标签集合, $B(s)$ 是商品 s 被打过的标签集合, $n_{u,b}$ 是顾客 u 打过标签 b 的次数, $n_{b,s}$ 是商品 s 被打过标签 b 的次数。如果从概率论的角度出发,认为顾客 u 喜欢商品 s 的概率取决于 u 曾经打过的标签,那么会得到如下的概率公式:

$$P(s | u) = \sum_b P(s | b) P(b | u) = \sum_b \frac{n_{b,s}}{n_b} \frac{n_{u,b}}{n_u}$$

对于标签 b , 令 $N(b)$ 为有标签 b 的商品的集合, $n(b, s)$ 为给商品 s 打上标签 b 的顾客数,可以通过如下的余弦相似度公式计算标签 b 和标签 b' 的相似度^[4]:

$$\text{sim}(b, b') = \frac{\sum_{s \in N(b) \cap N(b')} n_{b,s} n_{b',s}}{\sqrt{\sum_{s \in N(b)} n_{b,s}^2} \sqrt{\sum_{s \in N(b')} n_{b',s}^2}}$$

顾客对商品的兴趣通过如下的公式计算(本公式

认为顾客对商品的兴趣通过标签传递^[4]:

$$P(s|u) = \sum_b P(s|b)P(b|u) = \sum_b \frac{n_{b,s}}{n_b}$$

基于内容的商品推荐技术, 有着广泛的应用空间。对于一个电子商务平台来说, 有些顾客是登录之后进行浏览、购物; 对于更多的顾客来说, 并不进行登录, 而是随意的浏览。对于未登录而浏览的顾客来说, 电子商务平台应该记录他们的浏览轨迹, 并基于浏览内容进行分析, 在推荐位上及时进行推荐。对于非注册登录的顾客, 或者仅仅浏览的顾客, 现在很多电子商务平台采取的策略是: 通过顾客端的 cookie, 记录顾客的浏览轨迹并进行分析, 并据此向顾客推荐相关的商品。例如 YouTube (<http://www.youtube.com/>), 就采取上述推荐策略。这是一种基于内容的商品推荐技术。

2.2 基于协同过滤的推荐技术

协同过滤推荐技术, 通过分析顾客对商品信息的评价信息, 以发现顾客兴趣的相似性, 从而对具有相似购买兴趣的顾客进行交叉推荐。例如, 顾客甲喜欢物品 a, 顾客乙喜欢物品 a、b、c, 顾客丙喜欢 a 和 c。那么, 根据系统过滤的推荐方法, 因为喜欢物品 a 的人都喜欢 c, 所以给顾客甲推荐 c。协同过滤是基于大量历史数据集的, 因而存在稀疏问题和冷启动问题。在冷启动方面, 由于协同过滤是依靠人与人之间选择内容的相似度进行推荐的。因此, 与基于内容的方法相比, 不但存在新顾客问题, 而且还存在新对象问题, 即刚刚加入的对象如果没有被任何人选择过, 就很难被推荐^[2]。

基于余弦 (Cosine-based) 的相似度计算物品之间的相似度, 也就是说, 通过计算两个向量之间的夹角余弦值来计算商品之间的相似性。相关公式如下^[5]:

$$\text{sim}(i, j) = \cos(\vec{i}, \vec{j}) = \frac{\vec{i} \cdot \vec{j}}{\|\vec{i}\| * \|\vec{j}\|}$$

基于关联 (Correlation-based) 的相似度计算, 通过如下公式, 计算两个向量之间的 Pearson-r 关联度^[5]。

$$\text{sim}(i, j) = \frac{\sum_{u \in U} (R_{u,i} - \bar{R}_i)(R_{u,j} - \bar{R}_j)}{\sqrt{\sum_{u \in U} (R_{u,i} - \bar{R}_i)^2} \sqrt{\sum_{u \in U} (R_{u,j} - \bar{R}_j)^2}}$$

其中 $R_{u,i}$ 表示用户 u 对商品 i 的打分, \bar{R}_i 表示第 i 个物品打分的平均值。

根据之前算好的物品之间的相似度, 接下来对用户未打分的物品进行预测, 有两种预测方法: 加权求和; 回归。

加权求和的公式如下^[6]:

$$P_{u,i} = \frac{\sum_{\text{all similar items, } N} (S_{i,N} * R_{u,N})}{\sum_{\text{all similar items, } N} (|S_{i,N}|)}$$

其中, $S_{i,N}$ 为物品 i 与物品 N 的相似度, $R_{u,N}$ 为用户 u 对物品 N 的打分。

回归的公式如下所示^[3]:

$$R_N^i = \alpha \bar{R}_i + \beta + \varepsilon$$

物品 N 是物品 i 的相似物品, α 和 β 通过对物品 N 和 i 的打分向量进行线性回归计算得到, ε 为回归模型的误差。

2.3 推荐技术创新点研究

为了进一步改进推荐技术, 在现有技术的基础上, 本文提出了如下的创新点。

创新点 1: 商品推荐系统, 将商品分类, 变成一棵棵的树状结构。每类商品作为一个树状结构。推荐的时候, 如果顾客购买了一个树叶商品, 那么把树叶的 buddy 推荐给顾客。

创新点 2: 为了推荐系统, 必须查找不同商品的语义距离, 很多推荐系统使用了语义库 wordnet, 中国也有类似的语义库 Hownet。但通过语义库计算不同商品的语义距离是不合适的, 这是因为 wordnet 等语义库无法胜任这项工作, 应该建立针对电子商品平台的语义库。这是因为:

(1) wordnet 等语义词典, 是一个全领域的语义库, 很全面, 但是针对性不强, 并不是针对电子商务。

(2) 商品信息瞬息万变, 很多商品名字层出不穷, 很多商品在很短时间内消亡, wordnet 不可能包罗万象, 不可能及时有针对性的更新。例如: iPhone5 出现之后, 老旧的商品 iPhone2 可能面临下架, 面临消亡。

所以, wordnet 等全领域的语义库无法胜任电子商务的商品推荐系统, 为了及时更新, 为了增加准确性, 应该建立面向商品推荐系统的语义库, 这样便于准确查找语义类似的商品, 增加商品推荐系统的准确性。

创新点 3: 无论是物以类聚, 还是人以群分, 都有自身的问题, 所以引入知识推理机制, 建立推理规则库, 基于知识进行推荐, 基于关联规则进行推荐。也就是通过对顾客的购买行为进行推理, 准确捕捉顾客的心理, 从而精准的向顾客推荐商品。例如, 顾客购买了婴儿车, 购买了婴儿尿不湿, 购买了女性衣服、女性化妆品, 这个时候根据推理规则进行判断, 顾客购买了婴儿用品, 购买了成年女性用品, 这

名顾客是一名妈妈，婴儿车是长期用品，所以不向顾客推荐，尿不湿可以向顾客推荐，重点向顾客推荐其他婴幼儿用品，例如婴幼儿奶粉，以及刚生产的妈妈用品，例如妈妈食用补品。

创新点 4：通过以上对各类推荐方法的对比可以看出，每种推荐方法都有各自的优缺点，所以，单纯使用一种推荐技术往往是不可取的，在实际的电子商务推荐系统中应该采取多种技术相互结合、互为补充的混合推荐方法。

2.4 商品推荐的方式

商品推荐的方式主要包括两种：基于商品推荐位进行推荐；基于 EDM (Electronic Delivery Mail, 向顾客发送电子邮件) 进行商品推荐。所谓的商品推荐位，是指在顾客浏览、购买商品过程中，电子商务平台及时进行分析，在某一块位置，实时向顾客提供推荐商品的图片。商品推荐位是很多电子商务平台常用的商品推荐方式，如当当网 (<http://www.dangdang.com/>)、京东商城 (<http://www.jd.com/>)、亚马逊商城 (<http://www.amazon.com/>) 等。所谓的 EDM 商品推荐，是分析注册顾客的浏览、购买行为，然后通过电子邮件，向顾客推荐商品。采用 EDM 方式进行商品推荐的电子商务平台较少，如 sears (<http://www.sears.com/>) 等。很多电子商务平台，将基于商品推荐位的商品推荐方式和基于 EDM 的商品推荐方式结合起来，往往能够取得更好的商品推荐效果。

以 sears (www.sears.com) 为例，采取基于 EDM 的商品推荐策略。也就是分析注册顾客的浏览、购买行为，然后通过注册顾客的邮箱，发送推荐商品的信息。但是这种推荐方式，如果向顾客发送电子邮件过多，容易引起顾客的反感。而且，向顾客发送电子邮件进行推荐，必须获得顾客的电子邮件信息，对于那些浏览商品的顾客而言，通过发送电子邮件方式推荐是无效的。

下面是著名 B2C 电子商务平台京东商城 (<http://www.jd.com/>) 的主要商品推荐位介绍。其他 B2C 电子商务平台的商品推荐位与此类似。

(1) 首页上的推荐位

猜你喜欢(个性化推荐——基于浏览)推荐位

(2) 我的京东推荐位

猜你喜欢(个性化推荐——基于购买)推荐位

(3) 商品页面(图书/日百/3C)推荐位

最佳组合推荐位；

购买了该商品还购买推荐位；

浏览了该商品最终购买推荐位；

无货推荐推荐位。

(4) 过渡页与小推车页面推荐位

过渡页购买了该商品还购买推荐位；

小推车购买了该商品还购买推荐位。

(5) 三级列表上的推荐位

浏览了最终购买推荐位；

浏览还购买推荐位。

3 基于商品信息建立商品标签

建立基于商品信息的商品标签(中心词)，有两种方式：(1) 购物网站给商品做标签；(2) 客户对商品标签打分。下面分别加以论述。

3.1 购物网站给商品做标签

购物网站给商品做标签，是指商品上架之前，购物网站标注好商品的名称、规格、厂家、价格等信息。购物网站在每件商品上架之前，已经给商品做好了各种标签。

购物网站给商品做标签的时候，可以设置标签中每个量(商品关键词)的权重。但是从很多 B2C 电子商务网站的已有数据格式来看，并没有预先设定标签的权重。这很正常，商品数据格式是为了正确高效的描述商品信息，不会为某一项后续服务单独设置。

3.2 客户对商品标签打分(做标注)

客户以标签的形式，描述自己对商品的看法，对商品进行评价。让客户对商品做标签(标注)，就是让客户对商品信息进行描述和评价，实际上就是利用客户群体的智慧，广泛收集客户对商品的评价信息，增加商品推荐的针对性和有效性。

为了便于得到客户对商品的反馈，在客户浏览、购买某商品时，在商品旁边加上一个客户反馈条，让客户对商品进行打分，打分内容包括：商品品牌、商品价格、商品性能等。从而得到客户对商品进行评价的数据。客户对商品进行打分，实际上是客户对商品标签进行标注。

标签系统是客户兴趣与商品语义之间的桥梁。当客户对一个商品打上一个标签后，这个标签一方面描述了客户的兴趣，另一方面也表示了商品的语义，从而将客户和商品联系起来。

因此，标签数据是反应客户兴趣的重要数据源，而如何利用客户的标签数据来提高客户个性化推荐结果的质量，是推荐系统研究的重要问题。

一个客户标签行为的数据集, 一般由一个三元组的集合表示为 (u, i, b) : 客户 u 给商品 i 打上了标签 b 。当然, 客户的真实的标签行为数据远远比三元组表示的要复杂, 比如客户标注的时间、客户的属性数据、商品的属性数据等。

3.3 基于内容抽取商品标签(算法)

所谓的基于内容, 也就是基于客户的购买行为、浏览行为, 对商品的评价等。在这些数据的基础上, 学习客户的兴趣, 进行商品匹配, 向客户推荐相似的商品。

具体算法如下:

Step 1: 对客户的浏览行为设置权重, 购买行为设置权重。具体的权重应该经过统计得到;

Step 2: 根据客户对商品类的浏览次数, 以及客户购买商品类的次数, 计算客户对商品类的购买兴趣;

Step 3: 考虑客户在固定时间段内的购买兴趣;

Step 4: 分别计算客户对所有商品类的购买兴趣;

Step 5: 对客户的购买兴趣进行排序;

Step 6: 抽取前 n 名商品类, 向客户推荐。

4 判别商品的相似性、相同性

商品推荐, 必须进行商品的相似性判别。这里所说的商品相似, 包括如下几种情况:

(1) 语义相近。例如, “个人电脑”和“服务器”语义相近。

(2) 类别相同或者相近。例如, “蒙牛奶粉”和“三元奶粉”同属“奶粉”这一小类, 相似度高; “婴儿车”和“奶粉”同属“婴儿用品”这一大类, 相似度低。

4.1 商品相似度计算方法

(1) 语义相似度词典。例如, HowNet、WordNet 等语义相似度词典, 或者编制商品领域的语义相似度词典。

(2) 根据商品类别进行计算。对商品分类的过程, 实际就是给商品添加语义标签的过程。根据商品的类别, 计算商品的相似度, 是一个很好的办法。

4.2 相同商品判别方法

这里所说的商品相同, 是指语义相同。例如, “雅诗兰黛”和“Estee Lauder”是一回事, 语义相同。通过编制商品领域的同义词词典, 对商品相同性进行判别。

5 基于客户购物倾向的商品推荐

客户浏览某个商品, 系统基于浏览的商品信息即时推荐相关产品。

5.1 为什么基于客户购物倾向进行推荐

单纯基于某个商品信息进行商品推荐, 会有较大的盲目性, 因为每个商品有很多的属性, 包括品牌, 产品类别等等。系统无法判断客户感兴趣的属性。客户的历史行为、喜好等信息也无法反映客户本次购物的购物倾向。结合客户的购物倾向综合产品信息推荐, 使得推荐更加有针对性。

跟踪客户获取产品的路径, 可以得到客户的购物倾向。在网络购物的过程中, 客户通过电商的搜索引擎找到商品, 或者基于产品分类找到相关商品。客户在搜索引擎中输入的搜索信息, 或者客户基于产品分类逐步查找商品的过程, 都能很大程度反应客户的购物倾向。

从技术上来讲, 跟踪客户获取商品的路径, 分析客户购物倾向的关键词, 与产品信息关键词一起作为推荐相关产品的依据。

随着推荐层次的增加, 客户购物倾向的影响逐渐减弱, 在技术上通过降低描述倾向的关键词的权重实现。

假如客户购买兴趣没有变化, 比如在搜索引擎中输入的是“联想笔记本”→客户浏览了某款联想笔记本→在推荐的产品中客户浏览了推荐的某联想笔记本。则下次推荐的时候, 关键字“联想”和“笔记本”的权重进一步加大。则下一层推荐的一定跟“联想笔记本”相关的较多。客户购物倾向进一步影响推荐结果。

假如客户购买兴趣发生变化, 比如在搜索引擎中输入的是“联想笔记本”→客户浏览了某款联想笔记本→在推荐的产品中客户浏览了某 DELL 笔记本。则下次推荐时, 因为“联想”的权重已经降低, “DELL”的权重大于“联想”, “笔记本”的权重进一步增强, 此时推荐的产品跟“DELL 笔记本”相关的更多。客户初始倾向对推荐结果的影响减弱。

总之, 基于客户购物倾向的推荐能够结合客户本次浏览/购物的倾向为客户即时精准推荐相关产品。随着推荐层次的增加, 逐渐减弱客户初始倾向对推荐结果的影响, 动态适应客户兴趣的迁移。不会对推荐质量带来负面影响。

5.2 具体方案

(1) 识别客户的购物倾向

将客户购物行为看作一个前后无关的独立事件。根据客户在本次购物过程中的行为,根据商品种类,构建一棵商品分类树,并根据这棵商品分类树以及客户离散的购物兴趣点,对客户的购物倾向进行聚类,建立客户的购物倾向模型,分析识别客户的购物倾向。

可以通过多种方式,获取客户的购物倾向,无论客户如何找到当前商品,客户的查找方式、查找轨迹都在很大程度上反映了客户的倾向。例如,基于商品信息分类的多层次多维度导航、基于信息检索的商品特征匹配等。总之,客户可能通过多种渠道获取商品信息,那么跟踪客户获取商品信息的途径,就可以得到客户的购物倾向。

(2) 根据客户的购物倾向推荐商品

根据分析出来的客户购物倾向,计算出需要推荐的商品,并通过推荐位,向客户推荐。根据商品种类构建的商品分类树中,每个树叶结点和分支结点,都设置一个初始的关注值。如果客户对某个结点(树叶结点或者分支结点)表示关注,那么这个结点的关注值就相应的增加。当某个结点的关注值在短时间之内猛增的时候,说明客户对这个结点对应的商品类别,表现出极大地关注度和购物倾向,应该根据客户的购物倾向,将这个商品类别中的相关商品向客户进行推荐。

(3) 识别客户购物倾向的变化

关注客户购物倾向的转移,包括客户的购物倾向跳转和购物倾向偏离,并实时跟踪客户购物倾向转移的轨迹。如果客户购物倾向发生微小的偏离,那么降低客户购物倾向对推荐商品的影响权重;如果客户购物倾向发生大幅度的跳转,那么调整客户购物倾向相关度。也就是说,客户购物倾向模型动态自适应客户购物倾向的变化,并及时做出参数调整。

5.3 方案具体实施方式

Step 1: 构建树状的客户购物倾向模型

将基本的商品类作为本体,构建客户购物倾向模型的本体库。也就是说,从多层次的形式化模式上,给出商品类及其相互关系的明确定义,从而实现一定程度的知识共享和重用,并构建树状的客户购物倾向模型,提高系统互操作、可靠性的能力。

Step 2: 根据客户购物行为,识别客户的购物倾向

将客户本次购物行为,看作是一个前后无关的独立事件。根据客户的购物倾向模型,识别客户的购物

倾向。具体表现为如下几点:

- (1) 客户感兴趣的商品类别;
- (2) 客户的兴趣特征;
- (3) 客户对商品的访问频率;
- (4) 客户在商品上的逗留时间;
- (5) 客户对商品的收藏、检索结果反馈等情况。

Step 3: 跟踪客户的购物倾向轨迹变化

以时间为横轴(X轴),以符合客户购物倾向的商品所在商品类别为纵轴(Y轴),这样,就形成了二维平面。随着客户购物倾向的变化,这个二维平面上出现一系列的散点。根据这一系列的散点,进行客户购物倾向聚类,通过分析客户的购物倾向曲线的轨迹,能够跟踪客户购物倾向的变化。应该指出,客户的购物倾向可能发生偏移,也可能发生跳转。如果客户的购物倾向发生跳转,必须重构客户的购物倾向曲线。

Step 4: 预测客户下一步的购物倾向

在客户的购物倾向较为稳定时,也就是客户的购物曲线波动幅度较小的时候,预测客户下一步的购物倾向。如果客户的购物倾向突然发生跳转,那么重构购物倾向曲线之后,再预测客户下一步的购物倾向。总之,根据购物倾向的变化轨迹,修改客户购物倾向模型的倾向相关度,从而进行相关推荐。

Step 5: 确定商品推荐策略

根据客户的购物倾向,确定商品的推荐策略。

(1) 如果客户的购物倾向曲线较为平缓。客户的购物倾向曲线平缓,说明客户的购物倾向没有发生太大的变化,此时应该向客户推荐最相似的同类别商品。

(2) 如果客户的购物倾向曲线突然发生大幅偏转。如果客户的购物倾向曲线突然发生大幅偏转,说明客户的购物倾向发生了突然性的跳转。此时商品推荐应该随着客户购物倾向的变化,推荐符合客户新购物倾向类别的商品。

6 从 XBRL 的技术角度改进商品推荐方案

XBRL (eXtensible Business Reporting Language), 是可扩展商务语言的缩写。它是基于 XML (eXtensible Markup Language) 可扩展标记语言产生的、针对商务报告而设定的一种计算机通用语言,也就是说 XBRL 是 XML 中特定的一种语言。它有着与 XML 相同的可被各种软件和系统识别的功能,可以有效完成“商

业信息编制、分析和交流”, 达到低成本、高效率地提供准确可靠商业信息的功能。通过 XBRL 获取到的信息, 也无需打印或再次输入, 就可以方便快捷地运用于各种财务分析等领域。本文试图从 XBRL 的技术角度, 对商品推荐系统进行改进。目前只是做到了第一步: 从 XBRL 的角度, 在输入商品信息的时候, 将商品信息自动转换为 XBRL 格式。这样, 在做商品推荐系统的时候, 把推荐商品生成单独的 xml 页面, 页面内容包括商品标题、属性和价格等信息^[7]。

例如, 下面是一条用 XBRL 格式表示的书籍的信息。

```
<?xml version="1.0" encoding="gb2312"?>
<bookinfo>
<title>Core Java 2</title>
<author>Cay S.Horstmann Gary Cornell</author>
<publisher>POSTS TELECOM PRESS</publisher>
<price>80.00</price>
<recommendation>Fine<换行/>
                Good
                Read<换行/>
                Well
</recommendation>
</bookinfo>
```

这项工作有待进一步的改进。

7 结论和下一步的工作

本文首先介绍了商品推荐系统的相关背景, 然后

以京东商城为例, 归纳了电子商务平台的主要商品推荐位, 论述了如何基于商品信息建立商品标签, 以及如何判别商品的相似性、相同性。在上述工作的基础上, 重点论述了基于客户购物倾向的商品推荐方案。并从 XBRL 的技术角度, 对上述商品推荐方案进行了改进。

任何一种方案, 都需要进行实践的检验。目前为止, 本方案还没有在电子商务推荐系统中进行应用, 也就是说, 本方案还没有经过实践的检验。下一步的工作重点, 将把本方案应用于商品推荐系统中, 提高商品推荐系统的性能。

参 考 文 献

- [1] 许海玲, 吴潇, 李晓东, 等. 互联网推荐系统比较研究 [J]. 计算机学报, 2009, 20(2): 350-362.
- [2] 王宏宇. 商务推荐系统的设计研究 [D]. 中国科学技术大学, 2007.
- [3] 王树西. 基于内容的 B2C 电子商务推荐算法研究 [C] // 国际电子商务联合会中国分会第一届年会论文集, 2012: 78-85.
- [4] Beliakov G, Pradera A, Calvo, T. Aggregation Functions: A guide for practitioners. Springer [M]. Heidelberg: New York, 2007.
- [5] Torra V, Narukawa Y. Modeling Decisions. Information Fusion and Aggregation Operators [M]. Berlin: Springer, 2007.
- [6] Boritz J E, No W G. Assurance on XBRL-related documents: the case of united technologies corporation [J]. Journal of Information Systems, 2009.
- [7] Grabisch M, Marichal J L, Mesiar R, et al. Aggregation Functions [M]. Cambridge: Cambridge University Press, Encyclopedia of Mathematics and its Applications, 2009.